Assessing Army Values in New Soldiers:

Development of a Behaviorally Anchored Peer Evaluation Form

Stephanie T. Muraca

Initial Military Training Center of Excellence

Author Note

Stephanie T. Muraca-Grabowski, Research & Analysis Directorate, Office of the Deputy Commanding General of Initial Military Training.

Correspondence concerning this article should be addressed to Stephanie T. Muraca-Grabowski, U.S. Army Center for Initial Military Training, ATTN: ATCG-MTA, 210 Dillon Circle, Fort Esutis, VA 23604. Email: stephanie.t.muracagrabowski.civ@mail.mil

The views expressed in this manuscript are those of the author and do not necessarily represent the views of the U.S. Army or Department of Defense.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
03 FEB 2015	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Assessing Army Values in New Soldiers: Development of a Behaviorally Anchored Peer Evaluation Form	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Stephanie T. Muraca-Grabowski, PhD	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
United States Army Center for Initial Military Training, Research and Analysis Directorate, FT Eustis, VA	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this effort was to develop a standardized measure of Army Values inculcation in new Soldiers. The product is a peer evaluation form comprised of seven separate behaviorally anchored rating scales (one for each Value). This measure will enable Army leadership to take snapshots of individual IMT unitsâ values assimilation levels, to compare values assimilation levels over time or between units, and to evaluate the efficacy of competing Values training approaches. As an operational measure, the instrument could be used to identify deficiencies in values training while Soldiers are still in IMT, and while remediation is still possible.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**30**	

Abstract

The purpose of this effort was to develop a standardized measure of Army Values inculcation in new Soldiers. The product is a peer evaluation form comprised of seven separate behaviorally anchored rating scales (one for each Value). This measure will enable Army leadership to take snapshots of individual IMT units' values assimilation levels, to compare values assimilation levels over time or between units, and to evaluate the efficacy of competing Values training approaches. As an operational measure, the instrument could be used to identify deficiencies in values training while Soldiers are still in IMT, and while remediation is still possible.

Keywords: Army Basic Training, Army Values, peer evaluation, behavioral anchor, operational measurement

Assessing Army Values in New Soldiers:

Development of a Behaviorally Anchored Peer Evaluation Form

In the 1990's, the United States Army established a core set of seven values to guide the personal and professional behavior of its Soldiers. These values – Loyalty, Duty, Respect, Selfless Service, Honor, Integrity, and Personal Courage – are introduced in, and then reinforced throughout, initial entry training (IET),[1] and ultimately become the guiding force in a Soldier's endeavors (TRADOC Regulation 350-6; FM 6-22).[2] A Soldier who lives the Army Values is a professional committed to always "doing the right thing" in service to his/her country; an Army comprised of such Soldiers performs morally, ethically, and responsibly. Conversely, actions contrary to Army Values discredit not only the individual Soldier, but can erode unit cohesion, morale, and readiness, weaken Soldier resilience, cause others to question the Army's intentions or honor, and in extreme cases, endanger Soldiers' lives by galvanizing opposition forces.

Values assimilation is an essential component of civilian-to-Soldier transformation. Yet, this monumental undertaking must be accomplished by a time- and resource-constrained organization working to meet the operational demands of an Army at war. Because the cost of inadequate values training would be catastrophic, the purpose of this effort was to develop a standardized measure of Army Values inculcation in new Soldiers. The product is a peer evaluation form comprised of seven separate behaviorally anchored rating scales (one for each Army Value). This measure will enable Army leaders to take snapshots of individual IET units' values assimilation levels, to compare values assimilation levels over time or between units, and to evaluate the efficacy of competing Values training approaches. In light of this information, Army Values training practices and programs can be reaffirmed, readjusted, or replaced on a recurrent basis. Ongoing program review is the mechanism by which IET adapts to prepare an

ever-changing cohort of incoming Soldiers for the morally complex demands of the operational environment.

The Seven Army Values are not merely an artifact of the Army; to the contrary, they describe the deeply-held beliefs, traditions, struggles, and sacrifices of the American people (Army FY 01 Posture Statement, Chapter 5). In effect, the Army's values are America's values. What distinguishes the Soldier from the civilian is that the Soldier, in all professional *and* personal endeavors, is expected to *exemplify* loyalty, duty, respect, selfless service, honor, integrity, and personal courage (Army FM 1; TRADOC Pam 600-4). Concomitant with the Soldier's obligation to "live" the Army Values is his or her unqualified and unflagging subordination to the Constitution and civilian authority (Kohn, 2009). This confers legitimacy upon the Soldier's actions in defense of the nation and in support of the nation's geopolitical objectives. The Seven Army Values are described in Figure 1.

It is impossible to overstate the significance of the Army Values to the American Soldier, and it is impossible for those in the process of becoming Soldiers to avoid Values imagery, discussions, vignettes, videos, classroom lectures, and military-style briefings. Colorful and dynamic Army Values posters paper the walls in the barracks where IET Soldiers sleep, and in the classrooms and unit areas where they train and eat. All Soldiers are issued "dog tags" imprinted with the seven Values, and a "Blue Book" that describes contemporary heroes who exemplified the Values in combat. At the ranges where they learn to fire weapons and to move under fire, Soldiers are instructed to read and reflect upon monuments dedicated to fallen comrades whose final acts were a testament to the Values they embraced. However, even in this values-saturated environment, IET leadership does not rely on exposure alone to convey the values message to its new Soldiers. The Basic Combat Training Program of Instruction reserves

10.1 hours of total training time for formal classroom-based values instruction (BCT POI, 2010). The significance of values training becomes even more apparent when we compare its allotted 10.1 hours to the five hours reserved for such important classes as improvised explosive device (IED) detection and response.

Until recently, IET unit leaders developed and conducted their own Values training programs. Thus, is was conceivable that within a given training Battalion (approximately 1,200 Soldiers, or six Companies), one Company would teach Values via PowerPoint presentation, another would show clips from war movies and ask Soldiers to identify which characters did and did not display the Army Values, and yet another would have Drill Sergeants relate their own personal war stories and discuss how "doing the right thing" got them through difficult experiences. In the summer of 2010, Values training was standardized across IET. Now, while units are encouraged to develop and conduct their own supplemental training, the designated 10.1 hours of Values instruction is conveyed via a series of video vignettes produced by the United States Military Academy's Center for the Army Profession and Ethic (USMA-CAPE). Each vignette presents a situational conflict and then prompts Soldiers (via on-screen questions) to formulate a pro-values resolution during group discussion.

The Army Values message is pervasive throughout IET. However, incidents like the 2004 Abu Ghraib Prison Scandal and the 2006 Mahmoudiya rape-murder case, though anomalous and atypical, raise questions about whether that message is being received (Fay, 2004; Gebhardt, 2005; Kohn, 2009). Also causing concern is a chorus of social commentators who contend that today's Army-aged youth are, in every meaningful way, a less morally grounded generation than those that came before (most recently, Twenge, 2006; for summary and counter-point, see Trzesniewski & Donnellan, 2010). In this vein, Twenge and other critics argue that the "Me"

generation is marked by a keen sense of entitlement and an aversion to hard work and self-discipline. The veracity of this indictment, or even its applicability to those who distinguish themselves from their cohort by volunteering to serve, is beyond the scope of this paper. For our purposes, it is important to note that the indictment has resonated with IET leadership.[3] Consequently, with regards to Values inculcation, the question of whether or not "we're reaching these kids," has particular urgency.

Effort should be made to systematically measure Army Values inculcation in new Soldiers while they are still in IET, and while remediation, if necessary, is still possible. The Army will benefit from knowing if Values training programs are indeed reaching the next generation of Soldiers *before* potential disconnects manifest as prisoner-abuse scandals or murder-rapes on the evening news. The following sections outline the development of one such measure, and the methods used to assess its psychometric properties. This report concludes with a discussion of the measure's potential uses for operational assessment and research.

Behaviorally Anchored Rating Scales (BARS) and Peer Evaluation (PE)

Throughout the course of their initial entry training, new Soldiers must not only learn the Warrior Tasks and Battle Drills[4] that will enable them to fight and survive in combat, but they must also assimilate the seven Army values that will enable them to serve as Soldiers. Army Values inculcation is both a critical and complicated component of training – in a very short period of time, leaders must transform civilians, with conflicting civilian value systems, into Soldiers willing to live, and possibly die, by the same seven Army Values.

Competency in Warrior Tasks and Battle Drills is observable and quantifiable: leaders can watch a Soldier perform a task, and thus see if he or she has acquired a skill (in Army parlance, if that Soldier is a "go" or a "no go" on a given event). In contrast, values assimilation

is not a "go/no-go" event, and it is difficult for leaders to observe the extent to which their Soldiers live the Army Values. Leaders can get a *sense* of the effectiveness of values training by monitoring such things as unit indiscipline rates, cohesion, and attrition, but "a sense" is a far cry from systematic measurement, and cannot be used to gauge values assimilation over time, or to assess the effectiveness of one values training approach versus another.

The most straightforward way to systematically assess Army Values inculcation is to have each Soldier complete a questionnaire that asks him/her to rate his/her level of values assimilation on checklists or Likert scales. The problem with this face-valid, transparent approach is that frequently, individuals are inclined to provide the most socially desirable answer, and not necessarily the most honest one. Generally, impression management leads respondents to over-report good behavior/sentiment (here, values assimilation), and under-report bad behavior/sentiment, leading to inflated self-ratings and spurious relationships between variables (Fox et. al., 2007; Kline, 2005; Dovidio & Fazio, 1992; Knight, 1998; Devine, 1989; Nederhof, 1984; McBurney, 1994). Moreover, such a measure is likely biased by the tendency to conflate affect (e.g., "I feel that honesty is good.") with behavior (e.g., "I am honest."). To be effective, an Army Values measure must be able to distinguish between a Soldier who *lives* the Army Values and a Soldier who merely knows and/or agrees with them.

Toward the goal of developing an Army Values measure that captures behavior rather than affect or rote memorization, that minimizes social desirability bias, and that maximizes efficiency of administration and scoring, an Army Values (AV) Peer Evaluation (PE) instrument comprised of separate behaviorally-anchored rating scales (BARS) was developed (AVBARS-PE). Extant literature in BARS development and application demonstrates that *behavioral* anchors prime *actions* as the rating criteria for named performance dimensions (Smith &

Kendall, 1963; Hom et. al., 1982; Cardy & Dobbins, 1994; Cardy & Selvarajan, 2004; Phillips et. al., 2006). For our purposes, behavioral anchors prompt raters to consider how their peers behave, rather than speculate about what their peers know or believe, and ratings will reflect the extent to which Soldiers *act* in accordance with each Army value.

Peer evaluation has long been a valuable tool in school and workplace performance assessment, and has sparked military interest as a tool for leader assessment and development (Leigh et. al., 2007; Arnold, 2002; Dannefer et. al., 2005; Zaccaro et. al., 1999). However, peer evaluations are not immune to error. Raters may be moved by circumstance or motivated by ambition to intentionally distort evaluations of others (Wong and Kwong, 2007; Murphy & Cleveland, 1995 & 1991).[5] In this vein, raters may be inclined toward leniency and inflation in order to keep the peace among colleagues and promote workplace harmony (Murphy et. al., 2004). Conversely, a rater may assign unsuitably low scores to peers so that his or her performance appears exceptionally good by comparison (Gioia & Longenecker, 1994).

Research demonstrates that these biases can be controlled by managing raters' incentives and goals (Wong & Kwong, 2007; Cleveland & Murphy, 1992). Here, the approach is to deincentivize intentional distortion by making the entire AVBARS-PE process completely confidential, keeping individual raters out of the rating-feedback loop, and delaying feedback to raters' immediate supervisors. That is, Soldiers (raters) will neither have access to their individual scores nor to their unit's (i.e., Platoon's) aggregate score. Moreover, the Soldiers' immediate supervisors (i.e., Drill Sergeants and AIT Platoon Sergeants) will not have access to Platoon scores until after the rated Soldiers complete training, and leave the unit. Soldiers and supervisors will be made aware of this proviso prior to completing the AVBARS-PE, thus mitigating any rational incentive or pressure to artificially inflate or deflate peer ratings.

Immediately after the rated Soldiers depart the unit, supervisors score the instrument, aggregate ratings at the appropriate command level (i.e., Platoon, Company, Battalion, Brigade), review Soldier performance, and adjust training as necessary before training begins for a new "cycle" (or class).

Method

Development of AVBARS-PE

Smith and Kendall's (1963) seminal work on behaviorally anchored rating scales inspires many different approaches to BARS development. The approach used to develop the AVBARS-PE is Bernardin and Beatty's (1984) elegant, oft-employed four-step process (see Cardy & Selvarajan, 2004, for a useful blueprint): 1) identify performance dimensions; 2) develop dimension-specific critical incidents; 3) retranslate the critical incidents into the original performance dimensions; and 4) assess the performance of the critical incidents that survived retranslation.

1. Identify performance dimensions

The United States Army officially adopted the Seven Army Values in the late 1990's, thus identifying the performance dimensions, and completing step one. Throughout their career, enlisted Soldiers and Officers are continuously evaluated on the extent to which they know, understand, and exemplify the Army Values. Adhering to these values can affect Awards, promotions, selection to attend military schools, leadership positions, and desirable assignments (AR 623-3; DA Pam 623-3). So critical are the Army Values that – tactical, technical, and physical skills notwithstanding – Soldiers must demonstrate that they "understand, accept, and are prepared to live by the Values" before they can graduate from IET (TRADOC Pam 600-4).

The seven performance dimensions are: Loyalty, Duty, Respect, Selfless Service, Honor, Integrity, and Personal Courage.

2. Develop dimension-specific critical incidents

Following the logic of Bernardin and Beatty (1984), who used students to generate critical incidents for their performance dimensions, we used IET Soldiers. Specifically, we used IET Soldiers who had completed nine weeks of the 10-week BCT program, reasoning that Soldiers in this later stage of training will have sufficient exposure to Values training, and sufficient opportunity to develop not only behavioral patterns consistent with the Army Values, but an eye for recognizing behaviors consistent and inconsistent with Army Values as well.

Ninety IET Soldiers (45 male and 45 female) volunteered to participate in separate, confidential, hour-long focus group discussions (8 to 10 Soldiers per group) addressing the seven Army Values. The facilitator began each session by leading participants in a discussion about training in general, quickly steering them to the topic of Army Values. For each of the seven Values, the facilitator asked participants to define or describe the Value in their own words, discuss what that Value means to them, and then explain its purpose/usefulness (if any) in day-to-day Army life. After priming participants, the facilitator asked them to provide detailed examples of incidents in which they or other Soldiers exemplified the Value being discussed, examples of occasions in which they or other Soldiers acted in a manner contrary to the target Value, and examples of how an average Soldier might act relative to that Value. To avoid marginalizing Soldiers disinclined to participate in group discussion, the facilitator also provided participants with writing tablets and encouraged them to list/describe pro- and contra-Values incidents. This exercise generated a total of 106 critical incidents for the seven performance dimensions.

Notably, Soldiers had considerable difficulty explaining, describing, or identifying behaviors indicative of Loyalty and Honor. Because Honor is defined as, "living up to the Army Values," Soldiers were unable to conceptually disentangle it from the other Values to which it refers. Consequently, no unique incidents were generated for Honor. While Loyalty's definition clearly distinguishes it from the other Values, Soldiers seemed to wrestle with the conflict inherent in the concept – how to resolve competing loyalties. For example, Soldiers were uncertain about whether a behavior demonstrated loyalty or disloyalty if it involved being loyal to a peer at the expense of being loyal to a Drill Sergeant (e.g., failing to report a peer's wrongdoing to a Drill Sergeant). Consequently, relatively few unique incidents were generated for Loyalty. However, many Soldiers independently (in separate focus groups) volunteered that Loyalty had, "something to do with the Soldier's Creed." Figure 2 shows sample critical incidents for each of the Seven Army Values.

3. Retranslate the critical incidents into the original performance dimensions

The retranslation step is meant to ensure that incidents initially postulated as fitting a given dimension do, in fact, fit that dimension. A clean sample of 90 BCT Soldiers (45 male and 45 female) who completed nine weeks of a 10-week training cycle received a list of the 106 critical incidents generated in step two. The incidents were listed in random order, with no indication of the dimension, or Value, from which they had derived. Each Soldier was handed a list of the seven Army Values and their definitions, and then asked to match each critical incident with the Value he/she thought it best represented (or, in the case of the contra-Values incidents, with the Value he/she thought it most contrasted). Following Cardy and Selvarajan's (2004) recommendation, a minimum of 80 percent successful retranslation was the criterion for incident retention; an item was retained for AVBARS-PE only if 80 percent or more of the Soldiers

agreed that the incident represented the named Value. For example, 98% of Soldiers agreed that "sacrifices personal time to help other Soldiers" belonged in the "Selfless Service" dimension. Consequently, this incident was retained. Conversely, 47% of Soldiers agreed that "always stands up for his/her Battle Buddy" belonged in the "Loyalty" dimension, while 30% placed it in "Respect," 11% in "Selfless Service," 7% in "Integrity," and the remaining 5% in "Duty." Because there was no consensus as to where this incident belonged, it was rejected. This process resulted in the retention of 56 of the original 106 critical incidents.

4. Assess the performance of the critical incidents that survived retranslation

A group of 102 BCT Soldiers who had completed nine weeks of a ten-week training cycle, but who had not participated in either step two or three, evaluated the items retained after retranslation. Each Soldier was handed a list of retained incidents, and then asked to rate them on a 7-point scale, where 1 indicated that the incident reflected "extremely low/bad Army Values performance," and 7 indicated that the incident reflected "extremely high/good Army Values performance."

Here, in the final culling stage, we calculated means and standard deviations for each of the 56 incidents that had survived step three. An item's standard deviation was the critical test of whether or not it would appear in AVBARS-PE. A high standard deviation indicated that no consensus could be reached about the type of Values performance an incident represents (low/bad, moderate, or high/good), thus rendering it unsuitable for AVBARS-PE inclusion. Conversely, a relatively low standard deviation indicated that most Soldiers independently drew the same conclusion about the type of behavior an incident reflects. As such, items with low standard deviations were retained for AVBARS-PE. Using decision points established by Cardy

and Selvarajan (2004), items with standard deviations greater than 1.2 were rejected, and items with standard deviations below 1.2 were retained. This process left us with 39 viable incidents.

To create the AVBARS-PE instrument, we reunited each of the 39 incidents with its generating dimension (e.g., as determined in the retranslation stage, we returned "sacrifices personal time to help other Soldiers" to the Selfless Service dimension). We then used mean ratings to position the incidents on a 7-point performance continuum for their respective Values. Incidents with a mean rating of 2.99 or lower were situated on the "Low" end of their Value's performance continuum, incidents with a mean rating of 6 or higher were situated at the "High" end, and all other incidents (with mean ratings ranging from 3 to 5.99) were situated under "Moderate." Thus clustered, the incidents form behavioral anchors for each of the seven Values scales. Figure 3 shows two sample AVBARS-PE rating dimensions as they appear on the final instrument.

A product of Bernardin and Beatty's (1984) four-step process and Cardy & Selvarajan's (2004) BARS blueprint, AVBARS-PE is a soundly constructed instrument. The behavioral anchors (clustered critical incidents) should prime action as the evaluative criterion, and should maximize the reliability of evaluations within dimensions. That is, AVBARS-PE should capture the extent to which Soldiers act in accordance with the Army Values (vs. the extent to which they are perceived to agree with or know the Values) because, in effect, the instrument directs raters to consider and then align a peer's behavior along a continuum of identified low, moderate and high exemplars. Because the low, moderate, and high ranges of each rating dimension are clearly defined, each dimension should "mean" the same thing to different raters (e.g., each rater will not only be thinking about "Respect" the same way, but will have the same notion of what low, moderate, and high "Respect" looks like as well), thus yielding high inter-rater reliability.

The following sections will address these "shoulds," and evaluate the psychometric properties of AVBARS-PE. Specifically, we will examine mean dimension ratings for evidence of score inflation or deflation, estimate interrater reliability, and determined whether ratings were biased by ratee and/or rater gender.

Assessing the psychometric properties of AVBARS-PE

The AVBARS-PE was administered to 50 BCT Soldiers (25 male and 25 female) who had completed nine weeks of a 10-week training cycle.[6] A proctor instructed each Soldier to rate his/her Battle Buddy, and a selected Platoon member of the opposite sex. We designed the rating scheme so that each ratee was evaluated by exactly two raters – one male, and one female, and so that each rater rated exactly two peers – one male, and one female. Because of the constructed overlap, a total of 50 Soldiers were evaluated. During the informed consent process, the proctor insured participants that neither the Soldiers they rated, nor anybody in their Army chain of command (e.g., Drill Sergeants), would see or have access to the ratings, and that the ratings were for research purposes only. The proctor read AVBARS-PE instructions out loud to the Soldiers, and encouraged them to ask any questions they might have about the instrument, how to rate their peers, the purpose of the research, and the procedures in place to guarantee confidentiality. Total AVBARS-PE administration time, including informed consent, instructions and questions, was 23 minutes. Average time to complete AVBARS-PE (not including informed consent, instructions, or questions) was 11 minutes.

In his analysis of the performance rating process, Borman (1978) cautions that rating accuracy is attenuated when raters are assigned ratees with whom they have had only limited contact, and when raters have not had sufficient opportunity to observe their ratees' performance across the full range of named performance dimensions. Here, the unique insularity and

concentration of BCT allays this potential source of bias. BCT contains and controls Soldiers, training them in a "lock-down" environment that affords virtually no access to civilian society, television, computers, or to reading material other than religious texts, Army publications, or letters. Soldiers' time together is concentrated, constant, and intense. Consequently, they become keenly aware of each other's habits, behaviors and quirks. Battle Buddies[7] spend every waking hour together, and are usually assigned to sleep in bunked beds. In BCT, Platoon members stand together in formation, eat all of their meals together, and spend at least six days a week (Monday through Saturday), 15 hours a day, training together. On Sundays, following optional religious services, most Platoon members spend the day together engaged in maintenance, cleaning, and landscaping tasks. As such, we were confident that after nine weeks, each Soldier was sufficiently familiar with his/her same-sex Battle Buddy and an opposite-sex Platoon member to render informed evaluations.

Mean Ratings

Analysis of variance was conducted to estimate AVBARS-PE scale means, and to determine if observed differences between scale means were significant. Results and descriptive statistics are reported in Table 1. The distribution of scores around each mean suggests that raters considered the full range of each scale when assigning ratings. There does not appear to be a pattern of response setting at any point along the respective continuums of any of the rating scales. Each scale mean is situated just slightly above or slightly below the scale midpoint (four). This, along with the absence of counter-balancing response sets at the high and low ends of the scale and the relatively high standard deviations suggest that ratings were not distorted by range restriction (Murphy & Balzer, 1989), or leniency – severity (Hauenstein, 1992). Moreover,

significant differences between scale means suggest that raters were able to discriminate between the Army Values when assigning ratings.

Interrater reliability

Consistency between raters is key. If AVBARS-PE truly captures the extent to which a Soldier lives the Army Values, then, on a given dimension, a Soldier should receive the same rating from every peer rater. Any other outcome would suggest that AVBARS-PE is generating random error – perhaps reflecting one rater's affective orientation toward the Soldier, another rater's confusion about the behavioral anchors' meaning, and still another rater's individual interpretation of the Army Values. Percentage agreement is a straightforward way to assess consistency between raters. Unfortunately, many critics disagree with this approach because mathematically probable incidents of chance agreement may inflate, or, with truncated scales, hyper-inflate, the estimate of reliability (for a concise review, see Grayson and Rust, 2001). While the likelihood of chance agreement is inversely related to degrees of freedom, or number of potential scores the raters can assign, it is a pitfall that can be avoided by estimating Kappa rather than percent agreement. Cohen (1960) expresses Kappa as:

$$K = (p_a - p_c) / (1 - p_c),$$

where p_a is the observed proportion of agreements between raters, and p_c is the proportion of agreements expected to occur by chance (e.g., if raters score in a random manner). This more stringent test of interrater reliability deflates the effect of random chance agreement, and only credits exact agreement between raters (i.e., no credit is given to near misses or to series of ratings that are related linearly but not identically). Table 2 reports interrater reliability for each of the seven scales.

Using Landis and Koch's (1977) ranking of Kappa values from "poor agreement (K <0)" to "almost perfect agreement (K: 0.81 – 1.00)," we find that each of the seven Army Values scales – with the exception of Loyalty and Honor – fall within the "substantial agreement" range of 0.61 to 0.80. Consequently, we have evidence of high interrater reliability for Duty, Respect, Selfless Service, Integrity, and Personal Courage. Not surprisingly, given our initial Soldier samples' difficulty with Loyalty and Honor, the two scales evidence only "fair agreement (K: 0.21 – 0.40)." The "develop critical incidents" phase of AVBARS-PE construction revealed that Honor, as defined by Army Doctrine, is conceptually muddy (Honor is defined as "live up to the Army Values"). This muddiness makes it too slippery a gauge for Soldiers to wield with much precision. Loyalty, as a concept, is less murky, but the lingering "loyalty to whom" question minimizes its effectiveness as an evaluative criterion. Therefore, as expected, Loyalty and Honor perform poorly. Standard protocol in these circumstances is to omit the weaker scales from the final AVBARS-PE instrument. However, Soldiers are so accustomed to seeing *seven* Army Values, arranged in the same order (the acronym LDRSHIP), that to the initiated, the absence of "L" and "H" would be jarring. As such, we left Loyalty and Honor in the final AVBARS-PE, but acknowledge that ratings derived from those scales should be interpreted with caution.

The mixed-gender rater/ratee pairings allowed us to assess gender effects concomitantly with interrater reliability. Because each rating pair consisted of one male and one female rater, and that pair rated one male and one female peer, any gender distortion, be it same-gender favoritism, opposite-gender favoritism, or cross-gender favoritism or antipathy, would manifest as low Kappa values. In our examination, this was not the case. High Kappa values indicate that paired male and female raters independently agreed on ratings for their male and female peers. If gender bias was systemic, such agreement would not have been possible. Absent evidence of

gender bias in ratings, we have additional confidence that AVBARS-PE captures Army Values inculcation, and not Soldier popularity.

Discussion

AVBARS-PE shows promise as a measure of Army Values inculcation in new Soldiers. The results provide preliminary evidence of high reliability untainted by gender bias, range restriction, or leniency-severity. Because the assessment tool can be administered quickly (average 11 minutes to complete), it is appropriate for the fast-paced IET environment where efficiency is key and training necessarily trumps research and assessment. Clearly, AVBARS-PE falls short in measuring Honor and Loyalty. Arguably, given that the Army defines "Honor" as "living up to the Army values," and therefore makes it conceptually indistinguishable from those values, the measure of "honor" could be some combination of the other scale ratings – possibly one as simple as a grand or weighted mean. "Loyalty," however, is more vexing – it seems unlikely that we will be able to capture this construct until the Army clearly articulates a doctrinal resolution to the "potential competing loyalties" confound.

AVBARS-PE is a simple, straightforward instrument designed to be maximally accessible to IET Soldiers with regards to vocabulary, content, and response scaling. Moreover, a layperson can administer and score the instrument. Therefore, AVBARS-PE has potential operational applications at both the unit and individual levels. At the unit level, a "snapshot" of aggregate AVBARS-PE scores can provide leaders with insight as to how effectively the unit integrates Army Values into its training program. Additional snapshots over time can help unit leaders determine if they have remained on course, or if they should improve their values training approach. At the individual level, AVBARS-PE could be a valuable tool for Soldier professional development. For example, Cadre can administer the AVBARS-PE to Soldiers at

roughly the halfway point of BCT (five or six weeks into training). Leaders could provide feedback to Soldiers regarding their strengths and weaknesses, identify Soldiers who could benefit from additional attention or mentorship, and select Soldiers to serve as peer mentors. Soldiers could then be re-evaluated at the end of BCT (week 10); results could identify Soldiers who improved and Soldiers who could benefit from remediation prior to transitioning to AIT.

In applied research, AVBARS-PE could be a useful criterion measure for those interested in predicting Soldier outcomes based on Soldiers' baseline cognitive and non-cognitive attributes measured upon initial entry, and thus inform the development of selection and/or assignment tools. Those interested in performance enhancement could model end-of-training AVBARS-PE scores as a function of training intensity, training group composition and dynamics, and training leadership styles. Results could provide insight into how the training environment could be manipulated, or how training groups could be constructed, to maximize the likelihood and magnitude of values assimilation in new Soldiers.

References

Army Field Manual (FM) 1 (2005). The Army.

Army Field Manual (FM) 6-22 (1996). Army Leadership: Competent, Confident, and Agile.

Army Fiscal Year 2001 Posture Statement (2000).

Army Regulation (AR) 623-3 (2007). Personnel Evaluation: Evaluation Reporting System.

Arnold, L. (2002). Assessing professional behavior: Yesterday, today, and tomorrow. *Academic Medicine, 77,* 502-515.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77,* 975-985.

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent Publishing Co.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology, 63,*135-144.

Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: A consideration of alternative perspectives*. Cincinnati, OH: South-Western.

Cardy, R. L., & Selvarajan, T. T. (2004). Assessing ethical behavior: Development of a behaviorally anchored rating scale. Paper presented at the 2004 Midwest Academy of Management Meeting.

Cleveland, J. N., & Murphy, K. R. (1992). Analyzing performance appraisal as goal-directed behavior. In G. Ferris & K. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 10, pp. 121-85). Greenwich, CT: JAI Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Dannefer, E. F., Henson, L. C., & Bierer, S. B. (2005). Peer assessment of professional competence. *Medical Education, 39,* 713-722.

Department of the Army Pamphlet (DA Pam) 623-3 (2007). Personnel Evaluation: Evaluation Reporting System.

Devine, P. G. (1995). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56, 5-18.*

Dovidio, J.F., & Fazio, R.H. (1992). New technologies for direct and indirect assessment of attitudes. In J.M. Tanur (Ed.), *Questions about questions*. New York: Russell Sage Foundation.

Fox, S., Spector, P. E., Goh, A., & Bruursema, K. (2007). Does your coworker know what you're doing? Convergence of self- and peer-reports of counterproductive work behavior. *International Journal of Stress Management, 14*, 41-60.

Fay, G. R. (2004) Army Regulation (AR) 15-6. Investigation of the Abu Ghraib Detention Facility and 205th Military Intelligence Brigade.

Gebhardt, J.F. (2005). *The Road to Abu Ghraib: US Army Detainee Doctrine and Experience.* Fort Leavenworth, Kansas: Combat Studies Institute Press.

Gioia, D. A., & Longenecker, C. O. (1994). Delving into the dark side: The politics of executive appraisal. *Organizational Dynamics, 22*, 47-58.

Grayson, K, & Rust, R. (2001). Interrater Reliability. *Journal of Consumer Psychology, 10*, 71-73.

Hauenstein, N. M. A. (1992). An information-processing approach to leniency in performance judgments. *Journal of Applied Psychology, 90*, 842-856.

Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology, 67*, 568-576.

Kline, T.J.B. (2005). *Psychological Testing*. California: Sage Publications, Inc.

Kohn, R. H. (2009). Tarnished brass: Is the U.S. military profession in decline. *World Affairs, 171*, 73-83.

Knight, K. (1998). In their own words: Citizens' explanations of inequality between the races. In J. Hurwitz and M. Peffley (Eds.), *Perception and Prejudice.* Connecticut: Yale University Press.

Landis, J. R., & Koch, C. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Leigh, I. W., Bebeau, M. J., Nelson, P. D., Rubin, N. J., Smith, I. L., Lichtenberg, J. W., Portnoy, S., & Kaslow, N. J. (2007). Competency Assessment Models. *Professional Psychology: Research and Practice, 38,* 463-473.

McBurney, D. H., (1994) *Research Methods*. Pacific Grove, California: Brooks/Cole.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective*. Needham Heights, MA: Allyn & Bacon.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89,* 159-164.

Nederhof, A.J. (1984**).** Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15,* 263-280.

Phillips, J. K., Shafer, J., Ross, K. G., Cox, D. A., & Shadrick, S. B. (2006). *Behaviorally anchored rating scales for the assessment of tactical thinking mental* models. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith, P. C., & Kendall, L. M. (1963). Retranslation and expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149-155.

Twenge, J.M. (2006). *Generation Me: Why today's young Americans are more confident, assertive, entitled – and more miserable than ever before.* New York: Free Press.

TRADOC Pamphlet 600-4 (2010). The Soldier's Blue Book: The Guide for Initial Entry Training Soldiers.

TRADOC Regulation 350-6 (2011). Enlisted Initial Entry Training Policies and Administration.

Trzesniewski, K. H., and Donnellan, B. M. (2010). Rethinking "Generation Me": A study of cohort effects from 1976-2006. *Perspectives on Psychological Science, 5,* 58-75.

Zaccaro, S. J., Klimoski, R. J., Boyce, L. A., Chandler, C., Banks, D., & Gade, P. A. (1999). *developing a toolkit for the assessment of Army leadership processes and outcomes*, Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Footnotes

[1] Initial Entry Training is the beginning phase of the Army training process through which civilians are transformed into Soldiers. For most Soldiers, IET consists of two stages: 1) Basic Combat Training (BCT), where the Soldier is introduced to Army life, discipline, physical fitness, values, and the Warrior Ethos, and where he or she learns the basic tactical and technical skills necessary to perform in a hostile environment; and 2) Advanced Individual Training (AIT), where the Soldier learns the skills specific to his or her "job" or Military Occupational Specialty (MOS), while continuing to develop physically and in accordance with the Army Values and Warrior Ethos, and while maintaining common tactical and technical competencies. For some Soldiers, IET is completed in a single stage called One Station Unit Training (OSUT). Like their two-stage counterparts, OSUT Soldiers complete the standardized BCT program, but then remain with their BCT unit to learn the skills endemic to their MOS (Infantry, Combat Engineers, Military Police, Abrams Crewmen, Cavalry Scouts, and Chemical Biological Radiological and Nuclear Specialists are assigned to OSUT). Drill Sergeants lead BCT and OSUT Soldiers, while AIT Platoon Sergeants lead AIT Soldiers. IET unit structure (from most subordinate/smallest-to-highest ranking/largest) is: Squad, Platoon, Company, Battalion, Brigade.

[2] In 2005, the Army G-1 launched the Army Values Campaign Plan with the publication of the Executive Office Headquarters Army Values letter. The intent of the campaign was to, "re-emphasize the importance of Army Values throughout the force."

[3] Dr. Twenge was an invited speaker at a leadership development event at Fort Jackson, SC, the nation's largest gender-integrated Army Training Facility, in January 2010, and her *Generation Me* is on the TRADOC Pre-Command Course recommended reading list. The TRADOC Pre-

Command course is attended by all Officers and Noncommissioned Officers (NCOs) assigned to an IET Command position.

[4] Warrior Tasks and Battle Drills (WTBDs) are foundational tactical and skills common to *all* Soldiers, regardless of rank, assignment, branch, or occupational specialty. WTBDs include 15 tasks, 76 subtasks, and four battle drills, organized under the headings "shoot," "move," "communicate," and "survive and adapt" (TRADOC Regulation 350-6).

[5] Peer ratings may be distorted by (among other errors) range restriction (Murphy & Balzer, 1989), halo errors (Balzer & Sulsky, 1992), and leniency –severity (Hauenstein, 1992). Evidence of range restriction is found when ratings exhibit low variance (generally, low standard deviations) around an elevated (or depressed) mean rating, suggesting that ratings do not discriminate among different ratees based on their respective performance levels. Halo errors occur when a rater is disinclined or unable to distinguish among different rating dimensions for a single ratee, and is generally evidenced by high inter-dimension correlations. Leniency and severity is the tendency to assign an unsuitably high or low rating to a ratee.

[6] Raters were volunteer participants from one Company (220 Soldiers) of Soldiers assigned to BCT at Fort Jackson, South Carolina. Fort Jackson, the Army's largest gender-integrated Basic Training facility, trains an average of 65,000 Soldiers per year.

[7] TRADOC regulation 350-6 stipulates that throughout the entirety of their initial entry training, every Soldier must be accompanied by another Soldier of the same sex at all times (their Battle Buddy). Most Soldiers are assigned a Battle Buddy on the first day of BCT, and remain partnered with that same Battle Buddy until they graduate from BCT 10 weeks later. Under the battle buddy system, two Soldiers (the buddies) operate together as a single unit so that they are able to monitor, support, and assist each other as they acclimate to the fast-paced, high-stress,

and somewhat alien training environment. This regulation is enforced for the protection of individual Soldiers and cadre in AIT, BCT, and OSUT.

Table 1

*ANOVA and Descriptive Statistics AVBARS-PE**

				95% Confidence Interval	
	N	**Mean**	**Std. Deviation**	**lower**	**upper**
Loyalty	100	4.67	1.83	4.31	5.03
Duty	100	3.85	1.68	3.52	4.18
Respect	100	3.65	1.63	3.33	3.97
Selfless Service	100	4.49	1.56	4.19	4.80
Honor	100	3.98	1.74	3.63	4.33
Integrity	100	4.39	1.73	4.05	4.73
Personal Courage	100	4.51	1.63	4.19	4.83
Total	700	4.22	1.72	4.09	4.35

**F* (6, 693) = 5.29, *p* < .000

Table 2

AVBARS-PE Interrater Reliability

		95% Confidence Interval	
	K	**lower**	**upper**
Loyalty	.301*	.143	.459
Duty	.808*	.689	.927
Respect	.733*	.594	.872
Selfless Service	.778*	.647	.909
Honor	.383*	.219	.547
Integrity	.784*	.659	.909
Personal Courage	.807*	.686	.928

**p<.000*

Figure 1

The Seven Army Values

Value	Definition*
Loyalty	Bear true faith and allegiance to the U.S. Constitution, the Army, your unit, and other Soldiers.
Duty	Fulfill your obligations.
Respect	Treat people as they should be treated.
Selfless Service	Put the welfare of the Nation, the Army, and your subordinates before your own.
Honor	Live up to Army Values.
Integrity	Do what's right, legally and morally.
Personal Courage	Face and overcome fear, danger, or adversity (physical or moral).

NOTE: from TRADOC Pam 600-4

Figure 2

AVBARS-PE Sample Critical Incidents

	Low Value Performance	Average	High Value Performance
Loyalty	Leaves Battle Buddy hanging. Dimes out Battle Buddy to Drill Sergeants. Sees something wrong, but doesn't tell Drill Sergeants. Talks bad about the Drill Sergeants. Is totally unfamiliar with Soldier's Creed.	Stands up for Battle Buddy, but backs down if it makes him/her unpopular. Reports wrongdoing to Drill Sergeants if the Soldiers involved aren't his/her friends. Knows the Soldier's Creed, but doesn't always apply it.	Always stands up for his/her Battle Buddy. Will slow down during a run to help motivate Battle Buddy. Drill Sergeants can count on him/her to find out what's really going on in Platoon. Knows and lives the Soldier's Creed.
Duty	Oversleeps and misses fireguard shift. Sleeps during training. Drags feet during details. Late for formation. Loses gear; leaves canteen, headgear behind.	Sometimes needs a push to get started. May cut corners during training; looks for the easy way to get stuff done. Usually keeps track of his/her own gear.	Volunteers for details. Practices skills during free-time. Always ready to go with proper gear and knowledge. Will police-up room after a class and return things other Soldiers forgot.
Respect	Cops an attitude. Makes fun of Soldiers who are struggling. Talks down to peers. Talks back to/rolls eyes at Drill Sergeants. Stands at "parade pretty."	Sometimes lets ego get in the way of accepting help. Loses patience when other Soldiers are slow. Makes fun of peers behind their backs. Has some trouble with military bearing.	Shows proper courtesy to all. Always "locked up" in formation. Helps peers without putting them down. Avoids saying bad things about peers.
Selfless Service	Will only help out if he/she gets something in return. Shows-off for Drill Sergeants, but does nothing when they're not watching; spotlighter. Whines about things he/she misses from home.	Helps out when there's nothing better to do. Gets angry or upset if not rewarded/praised for effort. Talks a lot about wanting promotions or awards when they get out of IET.	Sacrifices personal time to help out peers. Is surprised when praised. Will slow down or re-do a task to help motivate a struggling Soldier. Talks a lot about loving the Army (vs. wanting promotions).
Honor	XXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXX
Integrity	Steals gear from peers or lets others get away with it. Talks/ moves around in formation if Drill Sergeant isn't watching. Takes food from DFAC. Tries to sell contraband.	Always behaves correctly when being watched, but lets behavior slide when Drill Sergeants aren't around. Doesn't try too hard during corrective action. Hopes nobody finds out if he/she does something wrong.	Never talks in formation or DFAC line. Tries to police-up other Soldiers if they mess up. Will take responsibility and do extra push-ups if he/she does something wrong.

Personal Courage	Hides in the back of the formation. Keeps going to the back of the line to avoid difficult training. Can't handle leadership positions, lets other Soldiers do the work for him/her.	Needs a lot of encouragement before trying something new or difficult. Has trouble handling student leadership responsibilities, something lets things slide instead of taking charge.	Will volunteer to be the first to try something. Will take charge of other Soldiers if they need guidance or discipline. Speaks up in class or training if he/she has a question.

NOTE: Colloquialisms and slang used by Soldiers were retained to preserve the realism of the critical incidents for their intended audience.

Figure 3

Sample AVBARS-PE Rating Dimensions

3. RESPECT: Rate each Soldier according to how well he/she Respects others.							
	Criticizes, laughs at, or talks down to peers. Gets defensive when others offer advice or help. Frequently interrupts or ignores others.		Sometimes cops an attitude with other Soldiers or Drill Sergeants. Ego sometimes gets in the way of accepting help. Seems impatient or arrogant when trying to help others.		Helps others without talking down to them; knows how to ask for/accept help when needed. Shows proper courtesy to all military and civilian personnel.	
	Low		**Moderate**			**High**	
Peer #1	1	2	3	4	5	6	7
Peer #2	1	2	3	4	5	6	7

4. SELFLESS SERVICE: Rate each Soldier's level of Selfless Service.							
	Only helps out if he/she knows they'll get something in return; always "looking out for number one." The first to quit, the last to volunteer.		Mostly helps out/volunteers when there's nothing better to do, or a Drill Sergeant is watching. Tends to whine if he/she does something right but doesn't get rewarded.		Sacrifices personal time to help other Soldiers; the first to volunteer for details. Is surprised if rewarded/praised for good work. Gives credit to the group rather than personally taking credit.	
	Low		**Moderate**			**High**	
Peer #1	1	2	3	4	5	6	7
Peer #2	1	2	3	4	5	6	7